DOCUMENT RESUME

ED 159 192 TM 007 407

AUTHOR Holley, Freda M.
TITLE Comparing Scores on the California Achievement Test (CAT) to Scores on the Sequential Test of Educational Progress (STEP).
INSTITUTION Austin Independent School District, Tex. Office of Research and Evaluation.
PUB DATE Jul 76
NOTE 14p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Academic Achievement; *Achievement Tests; *Comparative Analysis; Curriculum; Elementary Secondary Education; *Equated Scores; *Item Analysis; Mathematics; National Norms; Performance Factors; Reading Tests; *Testing Problems; Test Interpretation; Test Items; Test Validity
IDENTIFIERS Austin Independent School District TX; *California Achievement Tests; Minimum Competency Testing; *Sequential Tests of Educational Progress

ABSTRACT

To explain the discrepancy between median scores on the 1976 administration of the California Achievement Tests (CAT) and the Sequential Tests of Educational Progress (STEP) in the Austin Independent School District (AISD), ten technical variables typical of achievement tests were considered as explanations. (1) The STEP may measure different skills than the CAT. (2) Norm groups may differ. (3) The STEP may not measure what the high schools are teaching. (4) Curriculum sequencing of AISD high schools may not conform to that of the norm group. (5) Cross-level curriculum planning between elementary, junior high, and senior high levels may not be coordinated. (6) The AISD population may differ from the national population and hence from the norms. (7) Test familiarity may play a role in score differences. (8) The STEP is a more difficult test than the CAT. (9) The time of year of the administration may have depressed STEP scores. (10) Administration procedures differed from those used in the norming study. Of the ten variables considered, all but number 3 and possibly number 4 were accepted as possible explanations for the score differences. A further comparison was made between the CAT, STEP, and the CTB/McGraw Hill Proficiency and Review Tests for Reading and Numerical Proficiency. (CP)

rch And Evaluation

[redacted]

OFFICE
OF
RESEARCH AND EVALUATION

AUSTIN INDEPENDENT SCHOOL DISTRICT
TEXAS

austin independent school district

[redacted]

SUPERINTENDENT OF SCHOOLS

Dr. Jack L. Davidson

COORDINATOR, OFFICE OF RESEARCH AND EVALUATION

Dr. Freda M. Holley

Comparing Scores on the California Achievement Test (CAT) to Scores on the Sequential Test of Educational Progress (STEP)

July 1976

Freda M. Holley

Coordinator, Office of Research and Evaluation

Publication Number: 75-57

Comparing Scores on the California Achievement Test (CAT) to Scores on the Sequential Test of Educational Progress (STEP)

Achievement tests are not stable measuring instruments like meter sticks and thermometers. On the contrary they are what we might call "approximation instruments" because they measure the very difficult to capture construct called "human knowledge." Achievement test scores depend on two factors. The first is the human knowledge factor itself. If you have ever had the experience of being unable to remember the name of a person or thing you know quite well, the measurement problem associated with this factor will be easy to understand. One's knowledge is dependent on such things as one's emotional state, the setting in which the knowledge must be used or recalled, and even the time of day. In addition to this personal variability factor, however, achievement test scores are also subject to the technical variability through which the tests and the scores are derived. A meter measure, for example, can always be referred to one world standard maintained in France. Achievement tests have no such common referrent; they are dependent rather on a number of varying elements. Among these are: differing contents, differing norm group compositions, and differing levels of difficulty of the items of the test. These facts about achievement tests have to be taken into account when we look at disparate AISD median scores on different achievement tests. As the Office of Research and Evaluation (ORE) has considered the differences found this year on the two primary achievement tests we use which are referenced in the title, the following explanations have been considered.

1. The STEP may measure different things than the CAT.

   ORE finds this to be true. There is good evidence (see attachment 1) that the CAT is weighted toward the measurement of what we might call minimal basic skills while the STEP is measuring higher level academic competencies, Moreover, it may be that the possession of the minimal basic skill is a necessary, but not sufficient preparation for those higher level academic skills. Thus, a high score on the CAT would be necessary to achieve a high score on the STEP, but just because one had a high score on the CAT, he would not be guaranteed a high score on the STEP unless he also had much additional competency over and above that measured by the CAT.

2. The STEP norm group may be different from the CAT norm group.

ORE feels this also may be true. National norms are presumed to be representative of the national school population make-up. However, different companies define their own norm groups and there is no national standard for this. Thus, for example, one company may include private schools in their population, another may not. Also, test companies cannot force schools or students to participate in their norming group, and economics prevent their giving much economic reward for doing so. Therefore, norm groups rarely conform to precise sampling requirements necessary for true population representativeness. School systems thus suffer from the lack of a true national achievement standard.

There is some evidence that the STEP norm group and the CAT norm group are discrepant, based on evidence from the Anchor Test Study (a national study sponsored by the Office of Education that seeks to equate tests at certain grade levels.) For example, if we compare one level of the STEP (a lower grade level than the one AISD uses is the only one included in the Anchor Test Study) Reading test to the CAT eading Total, we consistently find a 4 to 5 percentile difference (see chart below). It is reasonable to expect that the same kind of difference will be found at the higher levels of the tests.

Predicted STEP Percentiles for the 50th Percentile of the CAT[1]

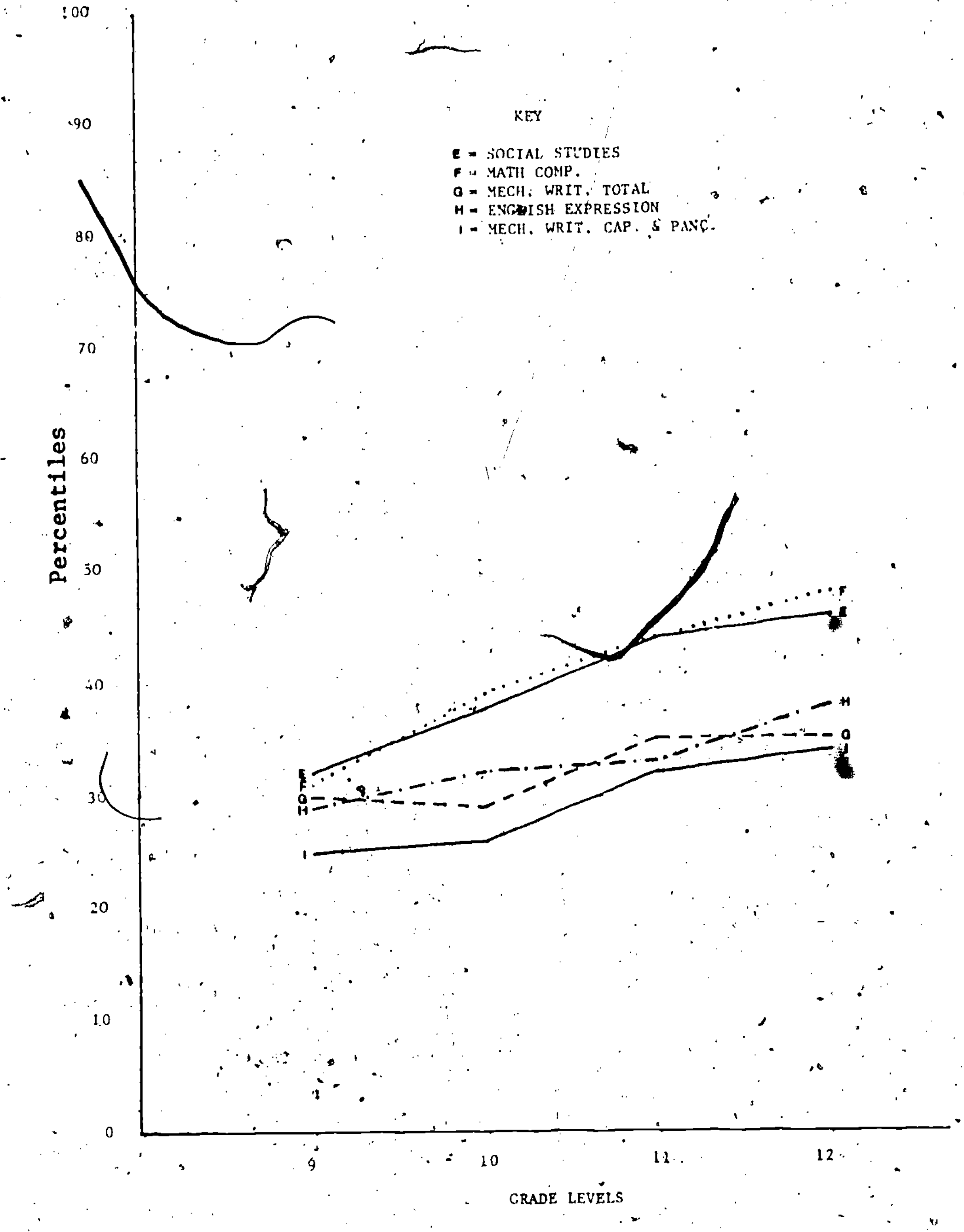
Grade	CAT Reading Total (%iles assuming March-June testing)				STEP Reading (%iles assuming April testing)		
4	50%ile (level 3)	=	Raw Score of 42	equiv. →	Raw Score of 31 (level 4)	=	46%ile
5	50%ile (level 3)	=	Raw Score of 53	equiv. →	Raw Score of 37 (level 4)	=	44%ile
6 6	50%ile (level 4)	=	Raw Score of 41	equiv. →	Raw Score of 42 (level 4)	=	47%ile
					Converted Scale Score of 435=		
					Raw Score of 27 (level 3)		

[1] Based on Table 1, page 9; Table 8, and Table 15 of the Anchor Test Study User's Manual - Equivalency and Norms Tables. Berkeley, California: Educational Testing Service, 1973.

One factor, the period of norming, would not appear to account for a discrepancy. Both tests were normed at approximately the same time, 1970. Both norms, incidentally, might now be considered out-of-date, and in view of national evidence of lower achievement this may result in Austin looking less well than it would were comparisons based on current national achievement score levels.

3. The STEP may not measure what the high schools are teaching.

ORE cannot accept this hypothesis for two reasons. First, the STEP was one of three test batteries selected by school and central office staff as being acceptable for AISD high school curriculum (see Figure B-1 in the <u>Systemwide Evaluation Technical Report 1975-76</u>). In addition, the upward movement of scores from 9 to 12th grade indicates in the graph below suggests a situation in which students are increasingly matching up to a curriculum.



4. The curriculum sequencing of AISD high schools may not match up to the norm group schools' sequencing.

ORE cannot adequately evaluate this hypothesis since the norm group school curriculum sequencing is unknown. However, one would expect national sampling to adjust for such difference. An example may serve to clarify this hypothesis. Say chemistry were nationally taught in the 9th grade and in Austin in the 12th grade. This would mean that AISD students would miss those chemistry items on the test until they reached the 12th grade. Some coordinators have expressed the feeling that this may be a factor and higher 12th grade scores may tend to confirm this as a possibility. However, one would expect that total scores would compensate for such a factor since all students receive the same test in all grades 9-12. That is, the student who had not yet had chemistry might get items say in Physics and thus compensate for the difference in sequencing.

5. The elementary, junior high, and senior high school curricula may not match up in AISD.

ORE feels that this also may be true. To some extent, of course, a perfect match would not be expected. Students begin to elect different scholastic pathways at the senior high level as they begin to prepare for future careers. However, the discrepancy between 9th and 12th grade scores and teacher comments about student preparation as they enter 9th grade suggests a discrepancy between high school entry expectation and earlier preparation. Moreover, there has traditionally been little cross-level curriculum planning that would lead to articulation between these school levels. Instructional coordinators and directors might well consider the possibility of this hypothesis.

6. AISD's population may differ from the national population and hence from the norm group population of the STEP or the CAT.

On the basis of vast national evidence that the composition of the school population on non-school factors will itself have an effect on achievement, it is to be expected that a school population make-up discrepant from the national make-up will affect percentile standings. Austin's school population differs in composition from the national population on a number of counts. To the degree that the test norms might biased toward national or AISD make-up, we might expect greater or less conformity on scores on the two tests.

7. Test familiarity may play a role in score differences.

AISD has been using the CAT for four years. Unconsciously even, personnel in AISD may have internalized the test content and have tailored instruction toward the curriculum content of the test. Also, students in grade 8 have taken the same test three years in a row. They too may be unconsciously learning toward the tests. This suggests percentiles at grade 6 should be closer to STEP percentiles than those at grade 8 as, indeed, is the case.

8. The STEP is a more difficult test than the CAT.

   This is comparable to saying that items 1 and 2 above are true. It does appear to ORE that the STEP is a more challenging test.

9. The time of year for the administration may have depressed STEP scores.

   This also may well be true. The only time in which the STEP could be scheduled in the 1975-76 school calendar was the week before and after Easter and no make-ups could be scheduled. This time could have affected both attendance at the test and student attitude toward the test. The CAT in grades 1-6 was given in April two weeks prior tc Easter and in grades 7-8 in February and make-ups were given.

10. Deviation of administration procedures from those used in the norming study.

   The CAT consists of only four subtests (2 reading, 2 math) while eight STEP subtests were given. In the STEP norming no more than 2 subtests were given per day; in the AISD administration, again for scheduling reasons, all 8 subtests were given in 2 days. The CAT was given over a 2 day period such that only the two subtests were given each day. Thus, fatigue may have acted to depress STEP scores. If 9th graders were assumed to be more easily subject to fatigue than seniors, the 9th to 12th grade upward movement of the scores would also tend to support this possibility.

The Content of Achievement Tests
in use in the
Austin Independent School District

There has been some thought given to the possible need for a "minimal skills proficiency" test for the Austin Independent School District (AISD). This led the Office of Research and Evaluation to prepare a comparison of the two current achievement tests used by AISD to one frequently used "minimal skills proficiency" test, the CTB/McGraw Hill Proficiency and Review Tests for Reading and Numerical Proficiency (popularly known as the Denver tests because they began as a test series designed for use in the Denver Public Schools). This comparison is of interest for two reasons. First, it appears that Level 4 of the California Achievement Test (CAT) now being used in AISD sixth to eighth grades is an adequate measure of the same thing tested by the proficiency test, particularly in mathematics. Moreover, it is evident that there is a great difference between the CAT and the Sequential Test of Educational Progress (STEP) with the STEP measuring skills and content at what could commonly be accepted as a much higher level than that of the CAT.

The two tables following contain the comparison first to the Proficiency and Review Test for Numerical Proficiency and second among the three reading tests.

Mathematics

Numerical Proficiency & Review Test Item	Item Description	CAT* Comparable Item	STEP Comparable Item
1	Add 2, 3-place numbers	2	
2	Add 2, 3-place numbers	2	
3	Add 4, 3-place numbers	3	
4.	Add fractions requiring a conversion (formula)	9 & 10	10
5.	Add money including hundreds-of-dollars and cents	1	
6	Add decimals (formula)	19 & 20	5
7.	Add ft. and inches	18	
8	Add mixed fractions	11 & 12	7 & 14
9	Add decimals (2 place plus 3 place)	19 & 20	
10	Add hrs. and mins. (3 sets)	-	
11	Subtract 3-place numbers	6 & 5	11
12	Subtract 3-place numbers	6 & 5	
13	Subtract money (formula)	14	1
14	Subtract 4-place numbers	7 & 8	
15	Subtract dollars + cents from tens of dollars + cents	13	
16	Subtract 2, 3-place numbers + mixed fractions	-	
17	Subtract 1-place decimal from 3-place decimal	23	26
18	Subtract mixed decimal + whole numbers	15,16,21,22	3 & 12
19	Subtract hours + mins. from hours only	24	
20	Subtract ft. + inches from ft. + in.(carrying necessary)	-	
21	Take dollars written out and show as figures	12	
22	Translate % to decimal	9	32
23	Translate % to fraction	-	34
24	Translate written numbers to figures	5	
25	Translate fraction to decimal	13 & 17	34 & 36
26	Translate written fraction to decimal figure	8	
27	Take percentage of money	11 & 15	
28	Take percentage of money	11 & 15	38
29	Find largest fraction	3	55
30	Find largest decimal	22	
31	Multiply 3-place number by 1-place number	25	
32	Multiply 3-place number by 2-place number	26 & 27	
33	Multiply decimals	42	
34	Multiply whole number by fraction	33	21
35	Multiply fraction by fraction	34 & 35	
36	Multiply mensuration (ft.xft., etc.)	41	30
37	Multiply 3 or 4-place number by 3-place	28	8
38	Multiply mensuration (ft.xft., etc.)	-	
39	Multiply decimals	43	
40	Multiply mixed fractions	36	4
41	Divide by 1-place number	31 & 29 30	

Mathematics Continued

Numerical Proficiency & Review Test Item	Item Description	CAT* Comparable Item	STEP Comparable Item
42	Divide by 3-place number	30	
43	Divide by decimal	32 & 39	18
44	Divide ft. or gallons	-	
45	Divide decimal	-	22
46	Divide decimal	-	
47	Divide ft. or gallons	-	
48	Divide by 2-place number	38	
49	Divide by fraction	47	29
50	Divide fraction by fraction	45, 46, 48, 49	

*CAT subtests are independently number and thus duplicate numbers here do not necessarily indicate duplicate items.

CAT and STEP Items beyond the Numerical Proficiency & Review Test Items

STEP

Computation: Items 6, 9, 13, 15-17, 19-20, 23-25, 27-20, 31-33, 35, 39-54, 56-60.

Examples (Patterned after test items, but not actual test items):

$$\frac{(1/3 + 1/3)}{(1/4 + 1/4) + (1/4 + 1/4)} = ?$$

The average of 6, 9, 7, 0, 4 and 1 is ?

If $r = 69$ and $d = 24$, then $\frac{rd}{3} = ?$

$$\sqrt{6} + \sqrt{21} = ?$$

Basic Concepts: Items 1- 50

Examples (Patterned after test items, but not actual test items):

If the area of a triangle is 64, then the area of the parallelogram (with a picture of a parallelogram in which the triangle is embedded) is ?

$(692)^2 = ?$

If $n-7 > 21$, then n can be ?

$c + d - (c+d)(c+d) = ?$

CAT and STEP Items beyond the Numerical Proficiency & Review Test Items — Continued

CAT

Computation: Items 4, 17, 44.

Example (Patterned after test item, but not actual item):

$3 \times (-4) = ?$

Concepts: Items 1, 2, 4, 6, 7, 9-11, 13-35.

Examples (Patterned after test items, but not actual items):

$\frac{4}{2,000}$ means the same as ?

means ?

Problems: Items 1-8, 10, 14

Examples (Patterned after test items, but not actual items):

One box weighs 10 pounds, another 12, and a third 17 pounds. What is their average weight?

John bought a refrigerator for $600. He paid $100 down and will pay the rest in 10 equal payments. How much will each payment be?

A triangle base is 10 inches; its height is 6 inches. What is its area?

Reading

Vocabulary

CAT	Reading Proficiency & Review	STEP
40 items	26 items	30 items
Items of comparable nature, some more difficult on CAT. All are multiple choice.		Identification of words, phrase, and sentences in context, difficulty may by higher.

Comprehension

CAT	Reading Proficiency & Review	STEP
4 reading selections (3 selections could be science or social studies 1 could be labeled as math) 10 items/selection + table of contents + index + diagrams	3 reading selections (All could be labeled science(health) or social studies) 8 items/selection	5 reading selections (1 selection science, 3 literary including drama dialogue) 5 to 8 items/selection